

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau



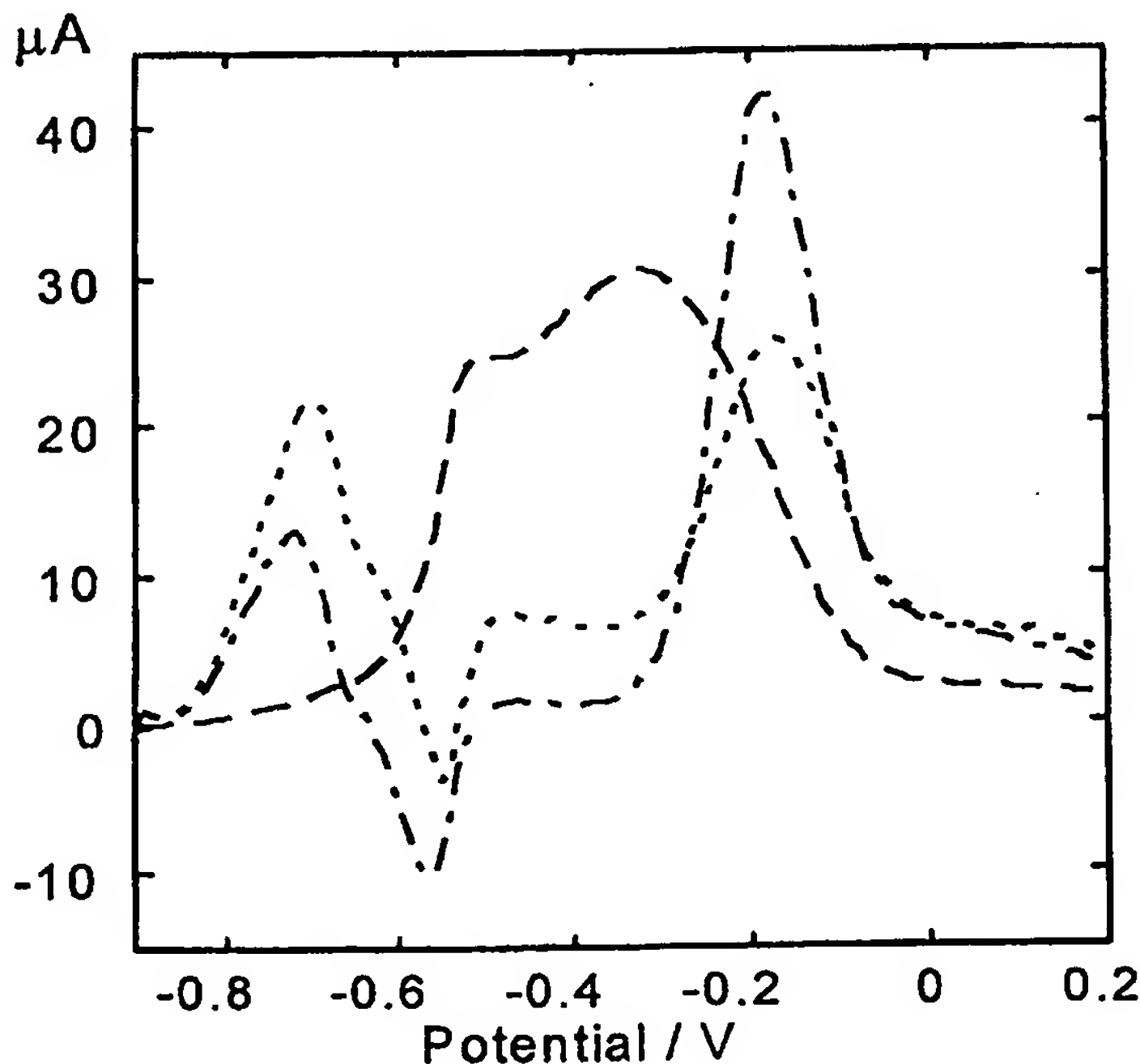
INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : G01N 27/416	A1	(11) International Publication Number: WO 00/20855 (43) International Publication Date: 13 April 2000 (13.04.00)
(21) International Application Number: PCT/GB99/03263 (22) International Filing Date: 4 October 1999 (04.10.99) (30) Priority Data: 9821482.8 3 October 1998 (03.10.98) GB (71) Applicant (for all designated States except US): CRANFIELD UNIVERSITY [GB/GB]; Cranfield, Bedfordshire MK43 0AL (GB). (72) Inventors; and (75) Inventors/Applicants (for US only): SAINI, Selwayan [GB/GB]; 16 Sandhill Close, Millbrook, Bedfordshire MK45 2JD (GB). BESSANT, Conrad [GB/GB]; "Apple-trees", 8 the Marsh, Carlton, Bedford MK43 7JU (GB). (74) Agents: STUART, Ian et al.; Mewburn Ellis, York House, 23 Kingsway, London WC2B 6HP (GB).		(81) Designated States: AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG). Published With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ANALYSIS OF MIXTURES

(57) Abstract

This invention relates to methods capable of measuring a large number of industrially relevant chemical compounds - the aliphatic organics. The invention employs an advanced electrochemical measurement in combination with enhanced information recovery techniques such as artificial neural networks (ANN). The resulting method achieves *virtual separation* (as opposed to physical separation in chromatography) and *simultaneous measurement* of individual aliphatic compounds in mixtures. The electrochemical determination suitably employs dual pulse staircase voltammetry. In this, cleaning pulses as used in conventional pulsed amperometric detection (PAD) are followed by a stepwise increasing potential, and current is measured after each step. The resulting signal can be analysed by ANN to determine the individual compounds in the mixture. An actual example shows the determination of individual concentrations of ethanol, fructose and glucose in mixtures of those compounds.



FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

- 1 -

ANALYSIS OF MIXTURESTechnical Field

The present invention relates to the analysis of mixtures. It particularly relates to the determination of organic compounds, especially aliphatic compounds, in mixtures.

Background Art

Measurement of the concentration of individual compounds in mixtures is usually achieved using some form of physical separation followed by a detection/measurement step. The most widely practiced instrumental technique is chromatography which, in its various guises, forms a cornerstone in chemical analysis.

Chromatographic instrumentation is routinely employed across a diverse range of industries, in medicine and in research and development applications. It is capable of providing high quality measurement information and it is used to set international standards in measurement and analysis.

However, chromatography does carry with it several less attractive features. It is a complex measurement technique requiring highly skilled personnel; it is usually confined to the laboratory (although a number of portable liquid/gas chromatography (LC/GC) systems are being reluctantly pushed into the field-analysis situation); it generally requires high capital and maintenance costs and introduces latency which can be significant where the sample is complex. Few systems are truly automated or standalone (chromatography columns need to be periodically cleaned or replaced etc.) which

- 2 -

would otherwise augment analysis in the industrial situation, particularly in remote monitoring applications. Nonetheless, these drawbacks are tolerated as viable alternatives are usually lacking in specific applications.

Chromatographic Separation and Measurement of Aliphatic Organic Compounds

Aliphatic organic compounds are a ubiquitous class of important industrial chemicals used throughout the food, petrochemical and pharmaceutical industries. Chromatography is comprehensively used in the separation and measurement of aliphatics when coupled to a suitable detector in liquid chromatography. Detection of aliphatic compounds can be achieved using one of several measurement techniques although in the last decade, a new electrochemical technique called pulsed amperometric detection (PAD) has emerged as the method of choice for the measurement of a large number of these compounds. Electrochemical PAD has proved popular because aliphatics generally show poor optical characteristics as there are no strong chromophores in their structure making UV-VIS detectors limited under routine conditions. Other detector systems such as optical density have proved useful in some determinations but generally are limited because of the indiscriminate nature of the detection.

PAD distinguishes itself from other LC electrochemical detectors because of its ability to substantially reduce or eliminate electrode fouling reactions that hampered previous attempts to determine

- 3 -

several aliphatic compounds. PAD eliminates fouling problems by using a solid metal electrode whose surface is electrochemically regenerated *in-situ* before each measurement without a change of solution [Johnson & LaCourse, 1990]. A number of commercial HPLC PAD instruments are available and used extensively in routine analysis. However, PAD can only be used in conjunction with chromatographic separation because the technique is not inherently specific. Therefore PAD detection following liquid chromatography separation is the current state-of-the-art in the measurement of individual aliphatic compounds in mixtures.

Disclosure of Invention

This invention significantly enhances the measurement of industrially relevant aliphatic compounds by removing the requirement for chromatographic separation prior to detection. In doing so, analysis is considerably augmented as the constraints that chromatography imposes in analytical measurements are likewise removed. The invention is capable of measuring the concentration of individual organic compounds, particularly aliphatic compounds, in mixtures. This makes possible a number of novel applications in industrial process monitoring.

Novel Methods for Measurement of Individual Aliphatic Compounds in Mixtures

Broadly, the present invention is based on a novel technique that employs a voltammetric technique, generally based on PAD, in combination with enhanced

- 4 -

information recovery techniques such as artificial neural networks. The resulting method has not been described previously and achieves *virtual separation* and *simultaneous measurement* of individual compounds in mixtures. The invention overcomes many of the drawbacks of chromatographic analysis in the measurement of aliphatic compounds.

According to the invention there is provided a method of analyzing a mixture of organic compounds comprising:

- (1) providing a sample solution containing a plurality of organic compounds to be determined;
- (2) immersing electrodes comprising a working electrode and at least one counter-electrode in said sample solution;
- (3) applying one or more potential pulses across said working electrode and at least one said counter electrode to regenerate the surface of the working electrode;
- (4) applying a varying potential across the working electrode and at least one said counter-electrode and measuring the electrochemical outcome, thereby providing an output signal related to the concentrations of the plurality of organic compounds in the sample solution; and
- (5) subjecting said output signal to chemometric analysis to extract respective data indicative of concentrations of each of said organic compounds.

Methods used in the invention

There are two main components to the novel method. The

- 5 -

first part is the physical measurement based on the electrochemical determination of (generally aliphatic) compounds in the mixture. This is followed by the virtual separation of the individual compounds and their quantitative measurement using multivariate calibration.

In order to utilise the electrode regeneration element of conventional PAD, its potential-time waveform has been appended to a stage of varying potential. However, unlike traditional PAD, the current is not sampled during the pulsed potential-time trace. Instead, it is sampled during the varying potential stage as in conventional linear sweep voltammetry. In the varying potential stage the potential may increase, usually essentially linearly, commonly in steps. Another possibility is a square wave potential as used in square wave voltammetry. When the waveform is stepped, the use of the combined waveform is referred to as *dual pulse staircase voltammetry* (DPSV). A similar waveform was described in the literature by Fung and Mo in 1995 but these authors did not develop the method further.

DPSV has two important consequences: it permits measurement of aliphatic compounds that are traditionally considered difficult to measure because of electrode fouling and significantly, yields a *spectra-like* response containing numerous peaks corresponding to different aliphatic compounds present in the solution. DPSV out-performs standard voltammetry for organic compound detection because of two cleaning pulses that are applied to regenerate the electrode before each scan. The first pulse is applied at a relatively high potential to form an oxide layer on the electrode surface which is then removed, along with any material adsorbed to the electrode, by a negative pulse [Roberts & Johnson, 1994]. The potential scan is then performed. To facilitate oxide

- 6 -

formation and dissolution, analysis is usually performed at a gold or platinum electrode in a solution of relatively low or high pH. The DPSV forms the preferred embodiment of the invention although other electrochemical methods might prove
5 favourable in certain applications.

The second part of the measurement substitutes conventional chromatography. In place of physical separation, the electrochemical response of the mixture is subjected to *virtual separation* using an artificial neural network (ANN) to
10 interpret the DPSV response. This enables the resolution and simultaneous quantification of aliphatic compounds in mixtures in the absence of chromatography. The motivation for using an artificial neural network to recover information from DPSV
15 voltammograms is many-fold. Information recovery from spectral traces (which are superficially similar to DPSV voltammograms) has been enhanced by the application of ANNs [Long et al., 1990; Schulze et al., 1995] and in a comparison with other multi-regression techniques such as partial least-squares regression (PLS) and principal components regression
20 (PCR), ANNs have been found to be superior for quantitative analysis of spectra [Goodacre et al., 1994]. Successful application of ANNs has also been reported for voltammetric data but, at the time of writing, the only papers in this area focus on easily resolvable and detectable metal ion species
25 with well documented electrochemistry [Cladera et al., 1997; Chan et al., 1997]. We have found that perturbation of DPSV responses can occur in mixtures as a result of inter-analyte interactions as well as variations in the effectiveness of electrode regeneration. Such features would appear to make
30 DPSV traces incompatible with most chemometric techniques, with the exception of ANNs which can effectively account for these variations by their ability to perform non-linear

- 7 -

mapping, generalisation and noise rejection. Therefore, in the preferred embodiment of the invention, artificial neural networks are used although recourse to other forms of chemometric operation is possible in certain applications.

5 A large array of different neural network architectures have been developed, but the virtually unanimous choice for spectroscopic and voltammetric work is a three layer feed forward network. In the case of DPSV interpretation, responses would be presented to the network's input layer.
10 This is connected to a hidden layer of non-linear neurons, which are in turn connected to a layer of linear output neurons, each of which would yield the concentration of one compound of interest in the mixture. Output neurons might also be included to provide additional information, such as
15 the likely error in the result. The ANN is trained on mixtures of known composition using the back propagation supervised learning algorithm, which builds the relationship between the response obtained and analyte concentration into the connections between the network layers. Training can take
20 some time (up to several hours) but once trained a network can interpret a response in under one second and its behavior is completely reproducible.

In order to show experimental evidence for the invention, methods that permit the measurement of individual aliphatic
25 compounds in mixtures is described herein as a series of experiments described below, with reference to the accompanying drawings.

Brief Description of the Drawings

30 *Figure 1: DPSV responses for 1mM glucose (-·-), 1mM fructose (---) and 32mM ethanol (—) at a platinum electrode in 0.1M NaOH. An example of a blank response for 0.1M NaOH is also shown (···).*

Figure 2: Blank subtracted DPSV responses for (a) 5.5, 11, 16.5 and 22mM ethanol, (b) 0.25, 0.5, 0.75 and 1mM fructose and (c) 0.25, 0.5, 0.75 and 1mM glucose. The arrows indicate the direction of correlation between increasing concentration and the peak heights.

5 *Figure 3: Blank-subtracted current contributions from glucose (---), fructose (—) and ethanol (—) at three key points in the DPSV responses for individual analytes.*

Figure 4: Blank subtracted current (in μA) at (a) -0.70V, (b) -0.32, (c) -0.23V in DPSV responses for 125 different combinations of glucose, fructose and ethanol.

10 *Figure 5: Example of the neural network architecture used to interpret DPSV responses. The input layer comprised 109 individual points (for clarity, only 23 are shown here). The number of hidden neurons ranged from 1 to 30.*

Figure 6: Relationship between network performance, number of hidden neurons and training time. Networks were testing using unseen voltammograms with analyte concentrations between those present in the training data.

15 *Figure 7: Relationship between actual concentration and concentration determined by a neural network with 30 hidden neurons, trained for 50,000 epochs, when tested on unseen data. Points marked by circles (O) are at concentrations present in the training data. Points marked by crosses (x) are at other concentrations. Error bars indicate one standard deviation. The diagonal lines indicate ideal response.*

20 *Figure 8: Relationship between actual concentration and concentration determined by neural networks with 5 (for ethanol and glucose) and 10 (fructose) hidden neurons, trained for 5,000 epochs, when tested on unseen data. Key as Figure 7.*

Figure 9: schematic view of apparatus for carrying out the method of the invention.

Modes for carrying out the Inventions

25 Simultaneous determination of sugars and alcohol in

- 9 -

mixtures

The mixture of aliphatics to be used to illustrate the invention comprises two closely related sugars, glucose and fructose, together with the alcohol substance, ethanol, in a common aqueous mixture. These compounds are also important detection targets in a number of industrial processes although their use here is mainly illustrative in order to show proof-of-principle rather than a specific application of the invention.

10 Materials and experimental procedures

Chemicals

Glucose and fructose were of analytical grade, supplied by Sigma (UK). Denatured ethanol was of HPLC grade (BDH, UK). Solutions were prepared in 0.1M NaOH produced by dissolving solid NaOH pellets (Fluka, UK) in reverse osmosis water.

Electrochemical Apparatus

Experiments were carried out in a static electrochemical cell 10, shown schematically in Fig 9, the design of which was dictated by the conditions required for DPSV. A 1.6mm diameter platinum disc working electrode 12 (BAS, USA) was used with a platinum wire counter 14 (BAS, USA) and a Ag/AgCl reference electrode 16 (Russell pH Ltd, UK). The cell contained 10ml of 0.1M NaOH to which small volumes (typically 20ml)

- 10 -

of analyte solution were added. A magnetic stirrer 18
(Rank Bros., UK) was used to return the cell to
homogeneity between scans by stirring at 500RPM for 10s.
Electrochemical measurements were performed using an
5 Autolab Pstat 10 (Eco-Chemi, Netherlands) and a
potentiostat 20 built by the authors and described
elsewhere¹⁴.

Electrochemical Procedures

Each DPSV scan consisted of a 3s 0.7V pulse, to
10 remove adsorbed fouling agents and form platinum oxide on
the electrode surface, and a 2s -0.9V pulse to regenerate
the surface by removing the oxide layer, followed by a
scan from -0.9V to 0.2V in steps of 10mV at a rate of
0.5V.s⁻¹. The current was recorded at the end of each
15 potential step during the scan. These electrochemical
parameters are based on those applied successfully
elsewhere^{12,13}.

Data Interpretation

Neural networks were implemented in Matlab 5 (The
20 Mathworks, USA) using the additional Neural Network
Toolbox. Matlab was executed in Windows 95 (Microsoft,
USA) on a 200MHz Pentium MMX based PC with 32Mb of RAM.
Using this system, training took between one and 20
minutes per thousand training epochs, depending on the

- 11 -

network architecture and amount of training data.

Interpretation of a voltammogram by any of the trained networks took a fraction of a second. For convenience, data interpretation was performed off-line, separately from the electrochemical experiments, although the two activities could be integrated on the same PC, or in a dedicated microcontroller, at a later date.

Both the input data and target matrices were scaled into ranges appropriate to the neuron transfer functions prior to training and testing. Each was range scaled between 0 and 1. The data was not mean centred. The actual concentrations are shown in Tables 1 and 2.

Table 1: Analyte concentrations used during network training

Normalised	Ethanol / mM	Fructose / mM	Glucose / mM
0.00	0	0.000	0.000
0.25	3	0.170	0.180
0.50	6	0.340	0.360
0.75	9	0.510	0.540
1.00	12	0.680	0.720

Table 2: Analyte concentrations used to test interpolation ability of trained networks.

Normalised	Ethanol / mM	Fructose / mM	Glucose / mM
0.125	1.5	0.085	0.090
0.375	4.5	0.255	0.270
0.625	7.5	0.425	0.450
0.875	10.5	0.595	0.630

- 12 -

Experimental Results and Discussion

The individual DPSV responses obtained for the three analytes are shown together in Figure 1, along with the blank response for NaOH. Clearly, the blank response is considerable. The traces shown represent the results of single scans, but they are typical of the responses obtained. The most variable aspect of the signal was found to be the blank response, possibly due to variations in the effectiveness of the electrochemical cleaning.

Individual Analytes

To clarify the characteristics of the individual analytes, blank responses obtained in NaOH only were subtracted from the analyte responses and the results plotted in Figure 2. From this it is clear that the key feature of the ethanol response is a wide peak centred around -0.32V while the sugars yield two peaks, one around -0.70V and another around -0.23V. However, each sugar makes a different contribution to these peaks, with fructose reacting mainly at -0.23V and glucose showing activity both here and at -0.70V. The bars superimposed on the responses in Figure 2 indicate the correlation between the analyte concentrations tested and the current at each point in the voltammogram. As expected, there is a very high positive correlation at the main peaks.

- 13 -

Fructose is peculiar in exhibiting a very strong negative correlation for the inverted peak at -0.56V.

The relationships between concentration and current are more clearly visualised in Figure 3, which shows the
5 blank subtracted currents at the centre of the three main peaks for a range of concentrations. The individual detection limits (according to the IUPAC definition¹⁵) for glucose, fructose and ethanol calculated from the data shown in Figure 3 are 40mM, 55mM and 2.2mM respectively.
10 The high limit for ethanol reflects the relative lack of sensitivity of DPSV to ethanol. Fructose has a higher limit than glucose because of variability in the fructose response which leads to poor correlation between concentration and current. The reason for this
15 variability is yet to be determined.

Mixture Analysis

Since each analyte has a distinctive response, it is conceivable that the voltammogram of a mixture of the three would produce a combined response from which these
20 responses could be resolved, thereby allowing determination of the concentration of each analyte. This requires that there are no prohibitive interactions between the analytes when they are mixed together and undergo analysis in NaOH. Such interactions would
25 include the complete suppression of an analyte response

- 14 -

by another analyte. To get an indication of whether this is the case, DPSV was carried out in 125 unique mixtures of glucose, fructose and ethanol with concentrations similar to those used when the analytes were studied individually.

Comprehension and visualisation of the resulting 125 voltammograms proved problematic due to the large amount of information involved. Although chemometrics can be used to reduce such data to more manageable representations, an appropriate chemometric technique could not be identified without first characterising the data. A method was therefore devised for visualising key elements of the DPSV responses: for each of the three key potential points identified earlier, blank subtracted currents were plotted as appropriately shaded tiles in a three dimensional analyte matrix. The results of applying this method are shown in Figure 4.

Considering the response at -0.70V (Figure 4a), the current clearly increases (the tiles become lighter) with increasing glucose. Furthermore, the current is reasonably uniform across all the results for each glucose concentration, indicating little interference from fructose and ethanol at this potential. This concurs with the individual results shown in Figure 2. At -0.32V (Figure 4b) the current generally increases

- 15 -

with increasing ethanol with little interference from glucose or fructose. Again, this corresponds to the results in Figure 2 but there are some notable features: the highest current (the lightest tile) at -0.32V is recorded for 12mM ethanol alone, but the presence of glucose or fructose suppresses this current considerably. Also, all the results for mixtures containing 3mM ethanol and 495mM fructose appear to be relatively high compared to neighbouring points - this seems to indicate erroneous results at that potential during that particular series of experiments. At -0.23V (Figure 4c) the current clearly increases with increasing fructose and also, to a lesser extent, with increasing ethanol. The effect of glucose at this point seems surprisingly small considering the large peak exhibited by glucose at this point in Figure 2, possibly indicating a resolution-enhancing interaction.

From this essentially qualitative assessment, it is clear that the mixed DPSV responses are not simple summations of the individual responses, indicating interference between the analytes. However, the responses have clearly been seen to contain characteristics of the individual analytes in proportion to their concentration. It should be noted that Figure 4 only provides three snapshots of the DPSV response,

- 16 -

presentation of all the voltammograms in this document would not be practical. The fact that the mixed responses are not simple summations of the individual voltammograms can be investigated more thoroughly using the latter method.

A significant clue to the nature of the inter-analyte interference is the fact that the currents measured in mixed solutions are lower than the sum of the individual responses. This is typical of fouling during measurement which could be caused by, for example, products of glucose oxidation at the first peak being adsorbed to the electrode and thereby reducing the surface area for the oxidation of ethanol and fructose at higher potentials. In addition, the calibration curves in Figure 3 showed signs of saturation at high concentrations, inferring that sensitivity will be reduced at the considerably higher concentrations created by the combination of analytes. Interference does not necessarily prohibit the use of DPSV as a way of quantifying the three compounds in a mixture, as long as the interference remains repeatable and can be successfully characterised. This could be achieved either through a detailed study of the oxidation, diffusion and fouling processes occurring at the electrode, or by multivariate calibration, as proposed

- 17 -

here.

Choice of Virtual Separation Techniques

Considering the observations described above, it appears that there is scope for virtual separation of the mixed responses if an appropriate chemometric methodology (possibly principal component regression (PCR)¹⁷, artificial neural networks (ANNs)¹⁸ or the Kalman filter¹⁹) can be identified. The resulting novel combination of DPSV and chemometrics should allow the resolution and simultaneous quantification of ethanol, fructose and glucose in mixtures without the need for a physical separation stage. Successful interpretation of voltammetric data by ANN²⁰⁻²², PCA²³⁻²⁵, the Kalman filter^{26,27} and PLS²⁸ has been reported. However, it seems likely that ANN methods will be most able to cope with the varying blank, non-linearity and inter-analyte interference observed in DPSV. In a comparison of multivariate regression techniques for the interpretation of spectroscopic responses, which are superficially similar to voltammograms, ANNs were found to yield more accurate results than PCR or PLS²⁹.

Neural Networks

Although many different types of ANN have been developed^{18, 30-31}, the vast majority of data interpretation

- 18 -

and reduction applications are most amenable to a multilayer feed forward network, so this type was used here. The operation and training of such a network is thoroughly documented elsewhere³². In short, a feed
5 forward network becomes, after sufficient training, a model able to map responses presented to it to a corresponding set of outputs. If sufficient training data is provided, the model is sufficiently generalised to map unseen responses to appropriate outputs. Once the
10 network is trained, it reduces to a simple matrix operation whose behaviour is fixed so it will always give exactly the same outputs for a given input pattern. There are two analytical functions to which feed forward neural networks are commonly applied: *classification* in
15 which the network is used to associate an input pattern with one of a finite number of possible targets (classification of odours using an array of non-specific gas sensors is a common example) and *multivariate calibration*, where the network maps input data to
20 continuous outputs (as used here). The outputs of a network when applied to unseen data are referred to as *predictions*.

In the present work an input is a 109 point DPSV response obtained from a mixture and the outputs are the
25 concentrations of the three analytes in the mixture. One

- 19 -

or more of the outputs may be zero, permitting the interpretation of binary mixtures, single analyte solutions and blanks. While the network type is common for many applications, the actual configuration of the network is acutely application-specific. The basic configuration used here (Figure 5) consisted of a linear input layer, a hidden layer of neurons with sigmoidal transfer functions and an output layer of three neurons (one for each analyte) with linear transfer functions. The sigmoidal hidden layer is critical as it allows the network to learn non-linear relationships between inputs and outputs (preliminary experiments with linear hidden neurons confirmed their unsuitability). The linear output layer reflects the expectation that the concentrations of the analytes are equally likely over a linear range.

Networks were trained using back propagation with adaptive learning rate and momentum, with random initial weights and biases. Random initial conditions were used to avoid picking fixed conditions which might favour one particular network design. This has the disadvantage of introducing variability into the performance of networks with exactly the same design, making it necessary to average the performance of a number of similar networks to get a true measure of performance for each design.

- 20 -

Training data was acquired by carrying out DPSV in known mixtures of the three analytes for the concentrations shown in Table 1. These concentrations were chosen because their DPSV responses are of comparable magnitude and this ratio of concentrations approximates the composition of alcoholic beverages and fermentation samples (possible application areas). Using these concentrations gives a total of 125 unique analyte combinations. The DPSV voltammograms for all these combinations were obtained experimentally and then combined in a 125 X 109 element input matrix. This was coupled with a 125 X 3 output matrix containing the corresponding analyte concentrations, to form what is termed a *data set*. Seven of these data sets were used to train ANNs, leaving the eighth free as a replicate for testing. To permit testing of the ANNs' abilities to interpolate between concentrations used in the training data, a further data set was acquired for all possible combinations (64 in total) of analyte concentrations between those in the other data sets, as shown Table 2. The interpolation set most accurately represents the calibration, so this was used for the majority of network testing.

Network Optimisation

The sole criterion for optimisation of the neural

- 21 -

network was achieving the smallest possible output error when applying the trained network to unseen DPSV responses. A qualitative assessment was provided by presenting a data set not used during training as a test set to the network and plotting actual concentrations against the corresponding network predictions. This was quantified for each analyte by calculating the root mean square (RMS) of the difference between the ANN predictions and the target concentrations. This essentially serves as a guide to the typical error, which is useful for comparing the performance of different networks.

For values distributed in a Gaussian way about their mean, the RMS equates to the standard deviation. Since the errors in the ANN predictions were found to exhibit an approximately Gaussian distribution, we can consider the RMS error to be the standard deviation about the true concentration for those points where the mean coincides with the true value. For a well trained network these points are in the majority. Hence, considering fundamental properties of standard deviation, we can say that (for a large enough sample) 68% of the NN predictions will be within one RMS error of the true value and 98% will be within twice the RMS error from the true value. Because of the large number of experiments

- 22 -

which had to be carried out manually it is not unreasonable to assume that the remaining 2% of points are outliers caused by experimental error. The maximum practical error can therefore be considered to be twice
5 the RMS error.

The RMS error was calculated using the normalised concentrations, to prevent the larger ethanol concentrations skewing the results. This has the secondary benefit that the RMS error can conveniently be
10 considered as a percentage of the full concentration range. Individual RMS errors were used for detailed studies of network performance, but were mean averaged over all three analytes to give a single value for a coarser, but more manageable, network comparison.

15 Optimisation of other network parameters was deemed unnecessary at this stage: the time required to train the network is unimportant as it only needs to be trained once, the fraction of a second taken to interpret a response is not worth minimising when compared to
20 measurement time, and the size of the network is irrelevant as the 30,000 or fewer bytes required to store the weights of a trained network are insignificant in modern computing terms. The parameters investigated to reduce the error were the training duration, the number
25 of neurons in the hidden layer and the number of training

- 23 -

sets. Often, data is pre-processed before being presented to an ANN but, since it was not clear what kind of pre-processing may be appropriate for DPSV, the data was used in raw form (except for linear scaling, described earlier). This makes the ANN's task more difficult as it has to cope with signal noise, variations in the blank response, irrelevant data points and erroneous responses, but it has the benefit of providing a simpler unitary interpretation system. It also prevents any biasing of the results caused by fortuitous pre-selection of pre-processing of the data.

Amount of Training Data

It is a widely held view that neural network accuracy increases with the amount of training data. This confirmed the interpretation of DPSV responses by comparing the accuracy of numerous networks of the same design which had been trained on varying amounts of data. The increase in accuracy was found to decay exponentially with respect to the number of data sets used for training, and no significant improvement was found when the number of data sets was increased from six to seven. The collection of additional training data was therefore considered unnecessary.

Hidden Neurons and Training Time

Networks were created with between one and 30 hidden

- 24 -

neurons and trained for up to 50,000 epochs. Each network was trained using the first seven data sets, and tested using the remaining set. For each network design three networks of that design were trained and tested and the average performance index calculated to reduce the variability caused by the use of random initial weights. Initially, different types of hidden layer transfer functions were evaluated. However, it quickly became clear that a *log-sigmoidal* transfer function (see Equation 1) gave the best results.

$$f(x) = \frac{1}{1 + e^{-x}} \text{ (Equation 1)}$$

The performance of networks with linear hidden neurons was consistently poor. Networks with hyperbolic tangent transfer functions often approached the performance of equivalent log-sigmoidal networks but frequently became trapped in local minima during training, especially for networks with a large number of hidden neurons.

Figure 6 shows the mean RMS error of networks for various combinations of hidden layer size and training time when tested on the interpolation set (the mean RMS errors of five similar networks were averaged for each column, to reduce variability induced by random initial weights). Although there is some variation, the general

- 25 -

trend is for the error to increase with training time.

Furthermore, networks with few hidden neurons are seen to be the most accurate. This is typical of overtraining, where the application of too many hidden neurons and too much training results in a network which is too sensitive to small anomalies in the training data, leading to poor results for unseen data.

The result of overtraining can be further seen by comparing the results obtained using a highly trained network with numerous hidden neurons (Figure 7) with a simpler network (Figure 8). As a method for determining the concentration of the three analytes over the complete range studied, the simpler networks are clearly most appropriate.

The protocol of individual analyte characterisation followed by acquisition and visualisation of mixture data, established here, can be used as a generic method of forecasting the possibility of analyte resolution in DPSV. This should help minimise the time necessary to build a picture of the generic applicability of DPSV by preventing unnecessary data collection and data analysis through identification of irresolvable mixtures at an early stage.

Experimental proof-of-principle has shown that a feed forward neural network can be used to determine the

- 26 -

concentration of ethanol, fructose and glucose in mixtures of those compounds in 0.1M NaOH, from a DPSV voltammogram acquired at an unmodified platinum electrode.

5 To obtain the best results for concentrations over the whole range it is preferable to use two relatively simple networks which have only been trained for 5000 epochs. These yield results with RMS errors of 5%, 6% and 6% of maximum concentration for ethanol, fructose and
10 glucose respectively. In the unlikely event that we are only interested in the concentrations used in training, more complex highly trained networks give more accurate results, reducing the RMS errors for fructose and glucose to 4% and 5% respectively. The fact that the errors are
15 not reduced by adding hidden neurons of training time suggests that these approximate to the lowest errors achievable with the current DPSV data. Much of this error could be caused by variations introduced by human error during the large number of repetitive experiments,
20 or by temperature variations. By training with responses for more unique concentrations it should be possible to harness the generality afforded by the simple networks and the high accuracy achieved by the more complex networks. Automation of the analyte mixing and data
25 acquisition would help as it allows much more training

- 27 -

data to be acquired in a given time and would enhance electrochemical repeatability by ensuring more accurate mixing times and sample additions.

For the simultaneous determination of glucose,
5 fructose and ethanol, the application of ANNs has transformed *state-of-the-art* DPSV from a mainly qualitative to a truly quantitative tool, enabling rapid measurement in mixtures of traditionally difficult to determine organic compounds without the need for a
10 physical separation stage, electrode replacement or complex electrode modifications. Other measurement techniques are available, but these do not have the simplicity of equipment and operation afforded by the combination of DPSV and ANN.

15 **Industrially applicable advantages of the invention**

The invention engenders a number of significant advantages in analytical measurements of aliphatic compounds. Novel advantages include orders of magnitude decreases in measurement time, vastly reduced instrument
20 and maintenance costs, automation and an improvement in the usefulness of information gleaned in measurements for certain applications. Further industrially applicable advantages are detailed below.

Generic applicability

25 PAD following LC has been carried out with a number

- 28 -

of different families of aliphatic compounds. These include aldehydes; carbohydrates (monosaccharides, oligosaccharides, polysaccharides); alcohols (aliphatic alcohols, polyalcohols); amino alcohols such as

5 alkanolamines; aminosugars and glycoconjugates; amines and amino acids; peptides and glycopeptides; sulphur compounds (e.g. thioalcohols, thioethers, thiophenes, thiocarbamates, organic thiosulphates). More recently, it has been shown that different penicillins can be

10 monitored using pulsed electrochemical detection. Many of these different compounds form important detection targets in process monitoring applications but are usually subject to the time and cost constraints associated with chromatography. Given the similarity of

15 the electrode regenerating process in PAD and DPSV, it follows that many of these compounds should be amenable to DPSV detection and virtual separation. Hence, the approach should have the potential to be used in the generic determination of aliphatic organic compounds in a

20 variety of industrial applications.

Speed of measurement and its advantage for rapid process monitoring

Firstly, the speed of the technique (typically less than 10s for a complete analysis) means that measurement

25 can be achieved in timescales that are orders of

- 29 -

magnitude less than conventional chromatographic analysis. This has implications in the application of the technique to on-line monitoring for process control and engineering. Whereas samples had previously
5 undergone the rigour of chromatographic detection with its compulsory time delays before a result is obtained, such measurements could essentially be real time. This facilitates examination of the process dynamics by evaluation of the concentration-time profiles of key
10 aliphatic compounds. This form of measurement is highly desirable in industrial process monitoring as it allows a higher degree of control over the process by taking steps based on monitoring to alter the conditions so that the system behaviour more closely follows a previously
15 optimised model.

Automation, usability and field measurements

The transference of separation from the physical to the mathematical domain makes the new technique ideally suited to automation and miniaturisation, allowing it to
20 be used in the field for rapid analysis by non-specialist personnel. Examples include field analysis of sugars in fruit, the determination of alcohol for the duty payable on alcoholic drinks (a portable monitor for customs and excise officers could be achieved). The nature of
25 electrochemical detection makes portability particularly

- 30 -

feasible - all that is required is a simple metal electrode in a flow cell and potentiostat circuit controlled and monitored by a personal computer or embedded microcontroller. This is in contrast to the complexity and cost of LC/GC detection systems, even where portability and dedication in an application has been achieved.

Instrument and running costs

The costs involved in producing an instrument are generally low, perhaps a fraction of the cost of a standard HPLC or GC. In addition, a possible instrument might incorporate two pumps, one which delivers sample and the other, the alkali or acidic electrolyte into a common flow cell in which the electrodes are sited. The ability to rapidly measure individual analytes in a complex organic mixture with simple, inexpensive instrumentation would be very attractive to many small industrial operations where expenditure on major equipment (such as LC/GC) is often limited. Such systems could be used as on-line monitors which could be totally self-standing and requiring very little maintenance. These might also be integrated to existing control systems.

Analytical parameters

In its present form, it is unlikely that sensitivity

- 31 -

enhancements (compared to chromatographic detection) would result in the instrument. Current detection limits for sugars are good - in the ppb range. However, there is further scope for improvement using alternative potential-time scanning procedures. For example, square wave voltammetry (SWV) which is known to significantly improve sensitivity in electrochemical measurements (by excluding capacitive noise) could be used in place of the current staircase voltammetry method used in DPSV. In addition, the SWV method achieves very good resolution of peaks in electrochemical measurements. This would improve the quality of measurement data presented to the ANN which is likely to benefit virtual separation and improve accuracy of measurement.

It is also noteworthy that the invention is likely to be able to out-perform chromatographic detection of some aliphatics. For example, chromatographic resolution becomes difficult when resolution of compounds of similar structure is required. These congeners are likely to show very similar chemical and physical behaviour reducing the opportunities for simple separation. A well known example is glucose and fructose in liquid chromatography. The results shown in Appendix A demonstrates that DPSV can clearly resolve the two related compounds as well as measure their individual

- 32 -

concentrations. Ethanol and methanol are further examples.

Literature cited:

1. Johnson, D.C.; LaCourse, W.R. *Anal. Chem.*, 1990, 62, 589A-597A.
- 5 2. LaCourse, W.R. *Analysis*, 1993, 24, 181-195.
3. Johnson, D.C.; LaCourse, W.R. *Electroanalysis*, 1992, 4, 367-380.
4. Rocklin, R.D.; Pohl, C.A. *J. Liquid Chromatography*, 1983, 6, 1577-1590.
- 10 5. Hughes, S.; Johnson, D.C. *Anal. Chim. Acta*, 1983, 149, 1-10.
6. Hughes, S.; Johnson, D.C. *Anal. Chim. Acta*, 1981, 132, 11-22.
7. LaCourse, W.R.; Owens, G.S. *Anal. Chim. Acta*, 1995, 207, 301-319.
8. Welch, L.E.; Mead, D.A.; Johnson, D.C. *Anal. Chim. Acta*, 1988, 15 204, 323-327.
9. Vandeberg, P.J.; Johnson, D.C. *Anal. Chim. Acta*, 1994, 290, 317-327.
10. Dobberpuhl, D.A.; Johnson, D.C. *J. Chromatogr. A*, 1995, 694, 391-398.
- 20 11. Altunata, S.; Earley, R.L.; Mossman, D.M.; Welch, L.E. *Talanta*, 1995, 42, 17-25.
12. Fung, Y. S.; Mo, S.Y. *Electroanalysis*, 1995, 7, 160-165.
13. Fung, Y.S.; Mo, S.Y. *Analyst*, 1996, 121, 369-372.
14. Bessant, C.; Saini, S. *Electroanalysis*, 1997, 9, 926-931.
- 25 15. IUPAC, *Spectrochimica Acta B*, 1978, 33, 242-245.
16. <http://www.cranfield.ac.uk/~c.m.bessant/data>
17. Massart, D.L.; Vandeginste, B.G.M.; Deming, S.N.; Michotte, Y.; Kaufman, L. *Chemometrics: A Textbook*; Elsevier Science Publishers B.V.: Amsterdam, The Netherlands, 1988.

- 33 -

18. Smits, J.R.M.; Melssen, W.J.; Buydens, L.M.C.; Kateman, G.
Chemometrics and Intelligent Lab. Systems, 1994, 22, 165-189.
19. Brown, S.D.; *Anal. Chim. Acta*, 1986, 181, 1-26.
20. Chan, H.; Butler, A.; Falck, D.M.; Freund, M.S. *Anal. Chem.*,
5 1997, 69, 2373-2378.
21. Cladera, A.; Alpízar, J.; Estela, J.M.; Cerdà, V.; Catasús, M.;
Lastres, E.; García, L. *Anal. Chim. Acta*, 1997, 350, 163-169.
22. Alpízar, J.; Cladera, A.; Cerdà, V.; Lastres, E.; García, L.;
Catásús, M. *Anal. Chim. Acta*, 1997, 340, 149-158.
- 10 23. Simeon, V.; Pavkovic, D.; Branica-Jurkovic, G. *Anal. Chim.*
Acta, 1992, 263, 37-42.
24. Díaz-Cruz, J.M.; Tauler, R.; Grabaric, B.S.; Esteban, M.;
Casassas, E. *J. Electroanal. Chem.*, 1995, 393, 7-16.
25. Grabaric, B.S.; Grabaric, Z.; Tauler, R.; Esteban, M.;
15 Casassas, E. *Anal. Chim. Acta*, 1997, 341, 105-120.
26. Brown, T. F.; Brown, S.D. *Anal. Chem.*, 1981, 53, 1410-1417.
27. Scolari, C. A.; Brown, S.D. *Anal. Chim. Acta*, 1984, 166, 253-
260.
28. Cabanillas, A. G.; Diaz, T. G.; Espinosa-Mansilla, A.; Lopez,
20 F. S. *Talanta*, 1994, 41, 1821-1832.
29. Goodacre, R.; Neal, M. J.; Kell, D.B. *Anal. Chem.*, 1994, 66,
1070-1085.
30. Melssen, W.J.; Smits, J.R.M.; Buydens, L.M.C.; Kateman, G.
Chemometrics and Intelligent Laboratory Systems, 1994, 23, 267-291.
- 25 31. Tusar, M.; Zupan, J.; Gasteiger, J. *J. Chim. Phys.*, 1992, 89,
1517-1529.
32. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. in Rumelhart,
D.E; McLelland, J. (Eds.), *Parallel data processing*. Vol. 1. MIT
Press: Cambridge, MA, 1986; pp318-362.

- 34 -

CLAIMS:

1. A method of analyzing a mixture of organic compounds comprising:

- 5 (1) providing a sample solution containing a plurality of organic compounds to be determined;
- (2) immersing electrodes comprising a working electrode and at least one counter-electrode in said sample solution;
- 10 (3) applying one or more potential pulses across said working electrode and at least one said counter electrode to regenerate the surface of the working electrode;
- (4) applying a varying potential across the working electrode and at least one said counter-electrode and
- 15 measuring the electrochemical outcome, thereby providing an output signal related to the concentrations of the plurality of organic compounds in the sample solution;
- and
- (5) subjecting said output signal to chemometric analysis
- 20 to extract respective data indicative of concentrations of each of said organic compounds.

2. A method according to claim 1 wherein the working electrode is a solid metal electrode.

25

- 35 -

3. A method according to claim 1 or claim 2 wherein the working electrode has a working surface of gold or platinum which contacts said sample solution in use.
- 5 4. A method according to any preceding claim wherein said at least one counter-electrode comprises a reference electrode.
- 10 5. A method according to any preceding claim wherein step (3) comprises applying a first positive pulse to said working electrode, and then applying a negative pulse to said working electrode.
- 15 6. A method according to any preceding claim wherein in step (4) an increasing potential is applied.
7. A method according to claim 6 wherein said potential increases stepwise.
- 20 8. A method according to claim 7 in which the current is recorded at the end of each step of said stepwise increase to provide said output signal.
- 25 9. A method according to any preceding claim in which the sample solution is stirred.

- 36 -

10. A method according to any preceding claim in which the chemometric analysis employs a technique selected from principal component regression, Kalman filter, PLS, or an artificial neural network ("ANN").

5

11. A method according to any preceding claim in which the chemometric analysis in step (5) employs an ANN.

10

12. A method according to claim 11 in which the ANN is a multilayer feed forward network.

15

13. A method according to claim 12 in which the ANN is a three layer network, having an input layer, a hidden layer of non-linear neurons, and a layer of linear output neurons comprising a respective output neuron for yielding data indicative of the concentration of a respective one of the compounds to be determined.

20

14. A method according to claim 13 wherein the hidden layer has sigmoidal transfer functions.

15. A method according to claim 14 wherein the sigmoidal transfer functions are log sigmoidal transfer functions.

25

16. A method according to any of claims 13 to 15 wherein

- 37 -

the hidden layer has ten or less hidden neurons.

17. A method according to any of claims 11 to 16 wherein
the ANN is trained on known mixtures using back
5 propagation supervised learning algorithm.

18. A method according to claim 17 wherein the ANN is
trained for less than 10,000 epochs.

10 19. Apparatus for carrying out the method of any
preceding claim having an electrochemical cell comprising
a vessel for holding the sample solution and electrodes
extending into the vessel; an electrical source for
applying potential to said electrode; and computing means
15 coupled to said electrode to receive the output signal
and to carry out chemometric analysis thereon.

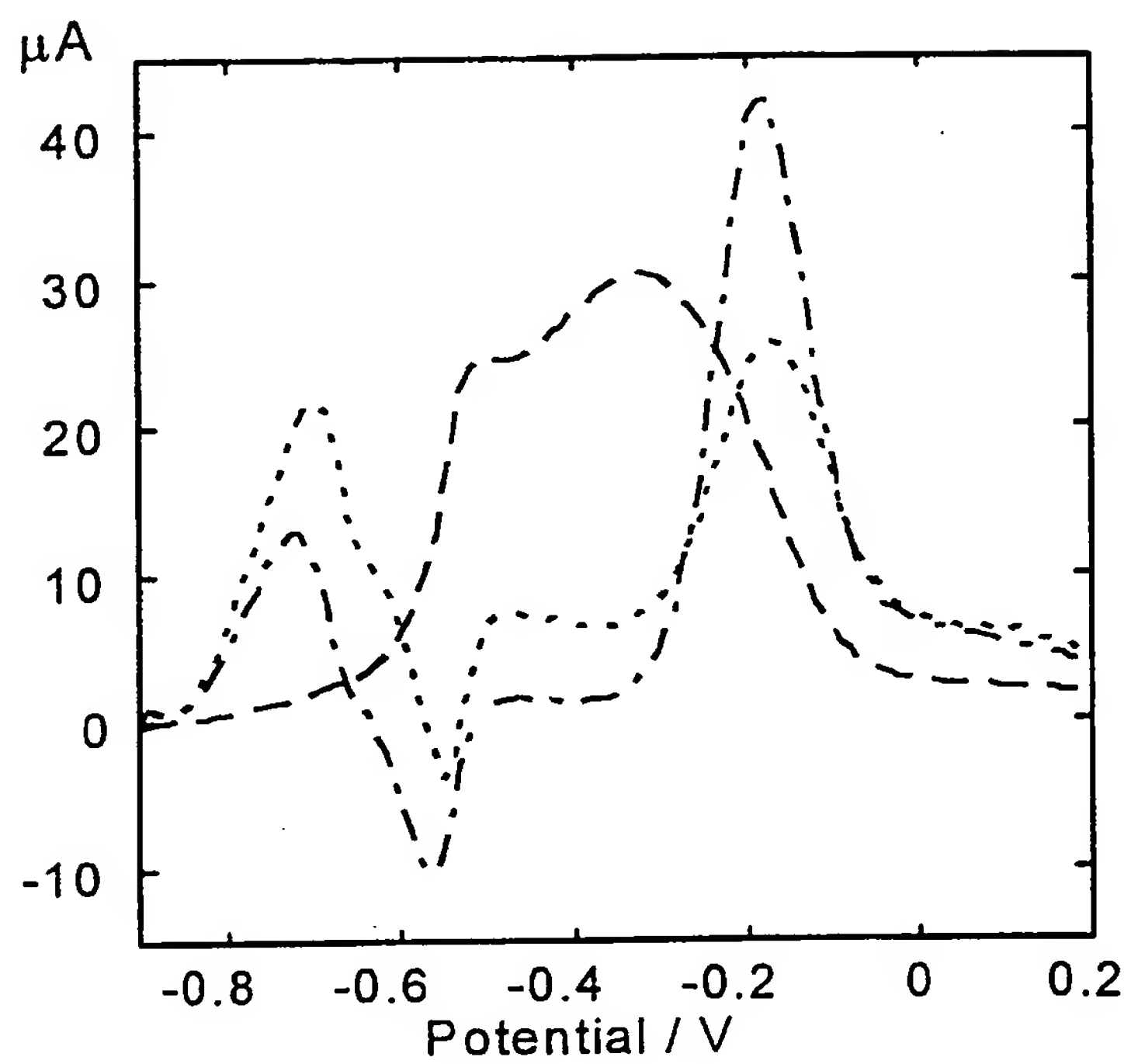
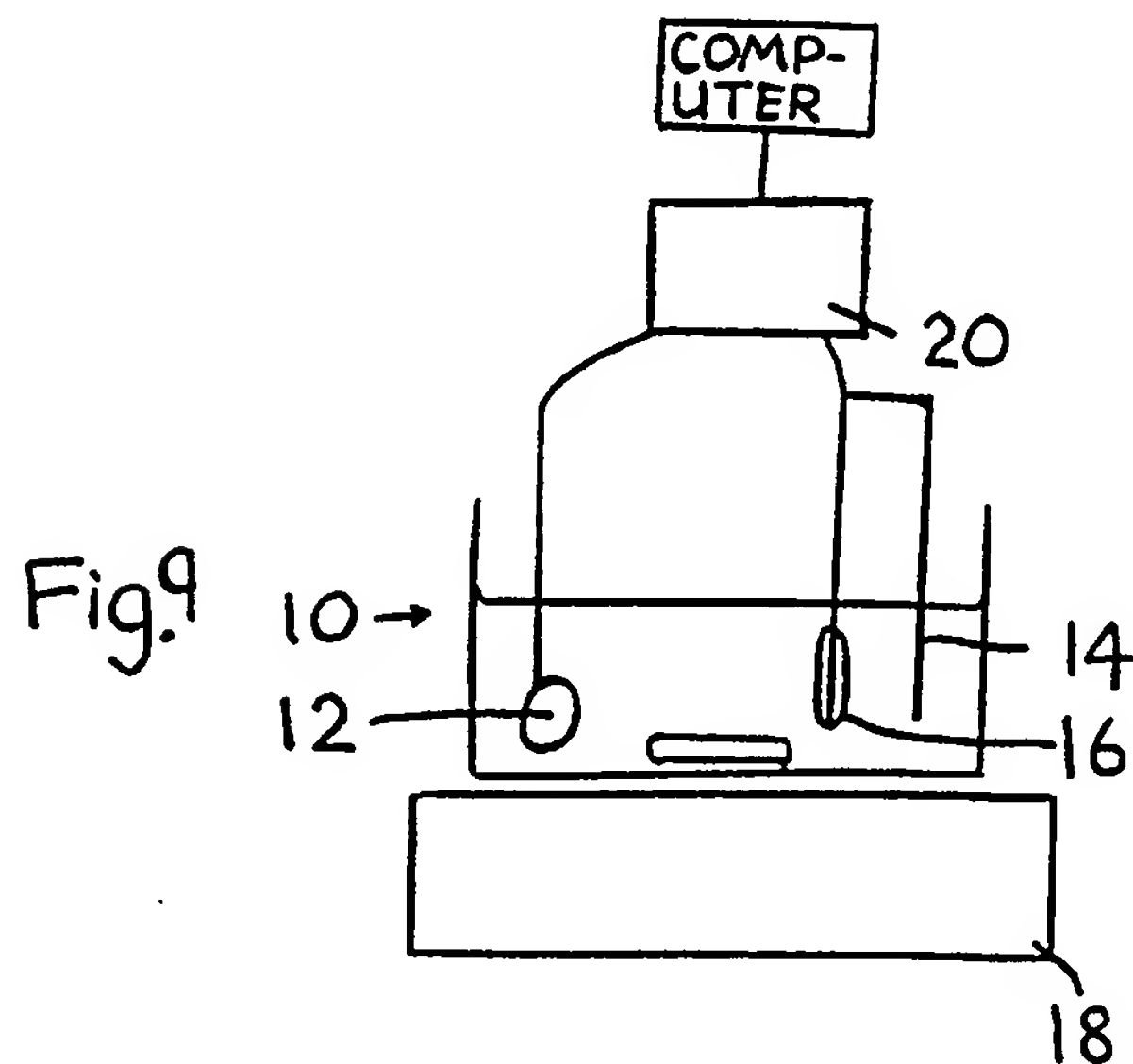


Fig 1



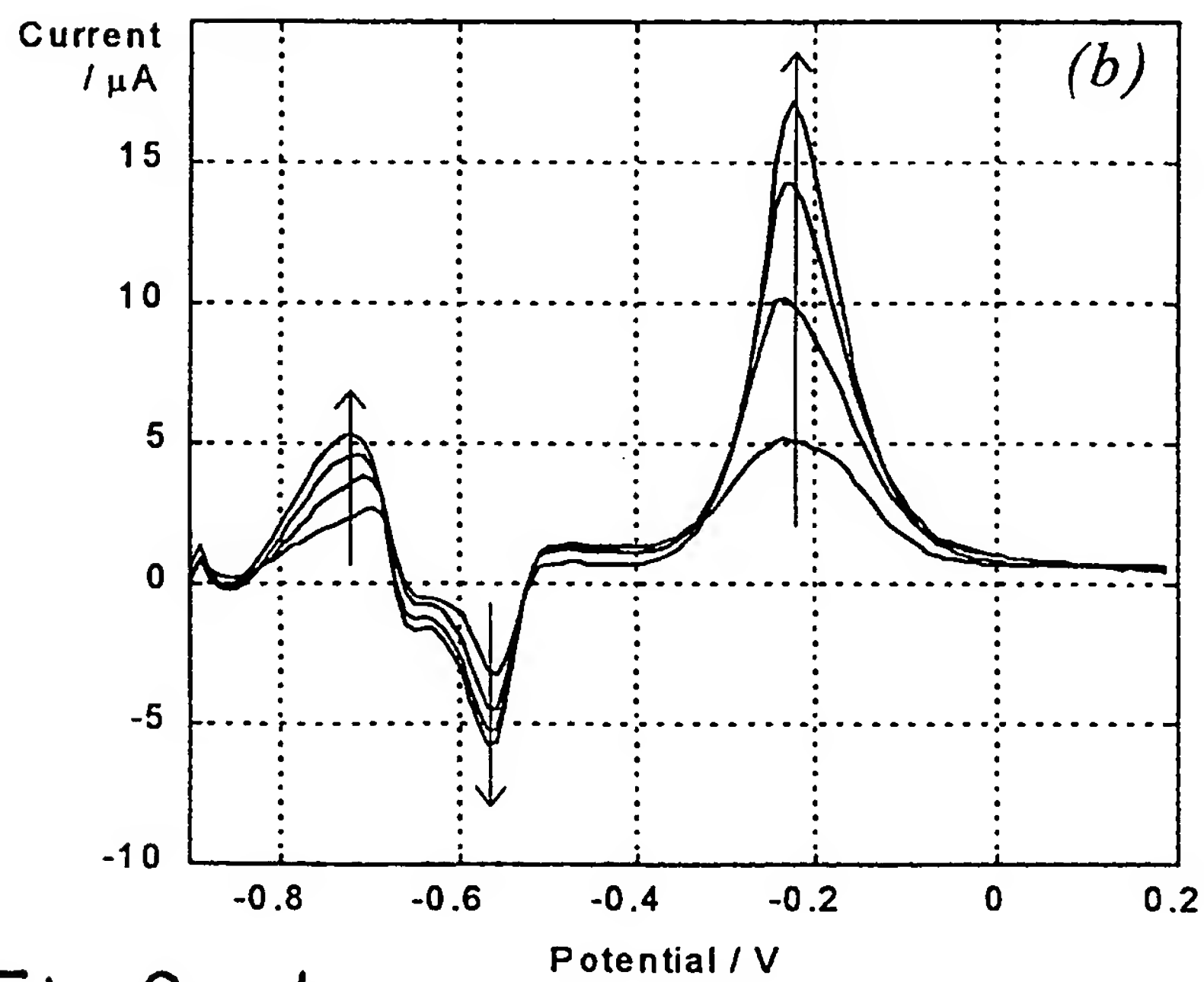
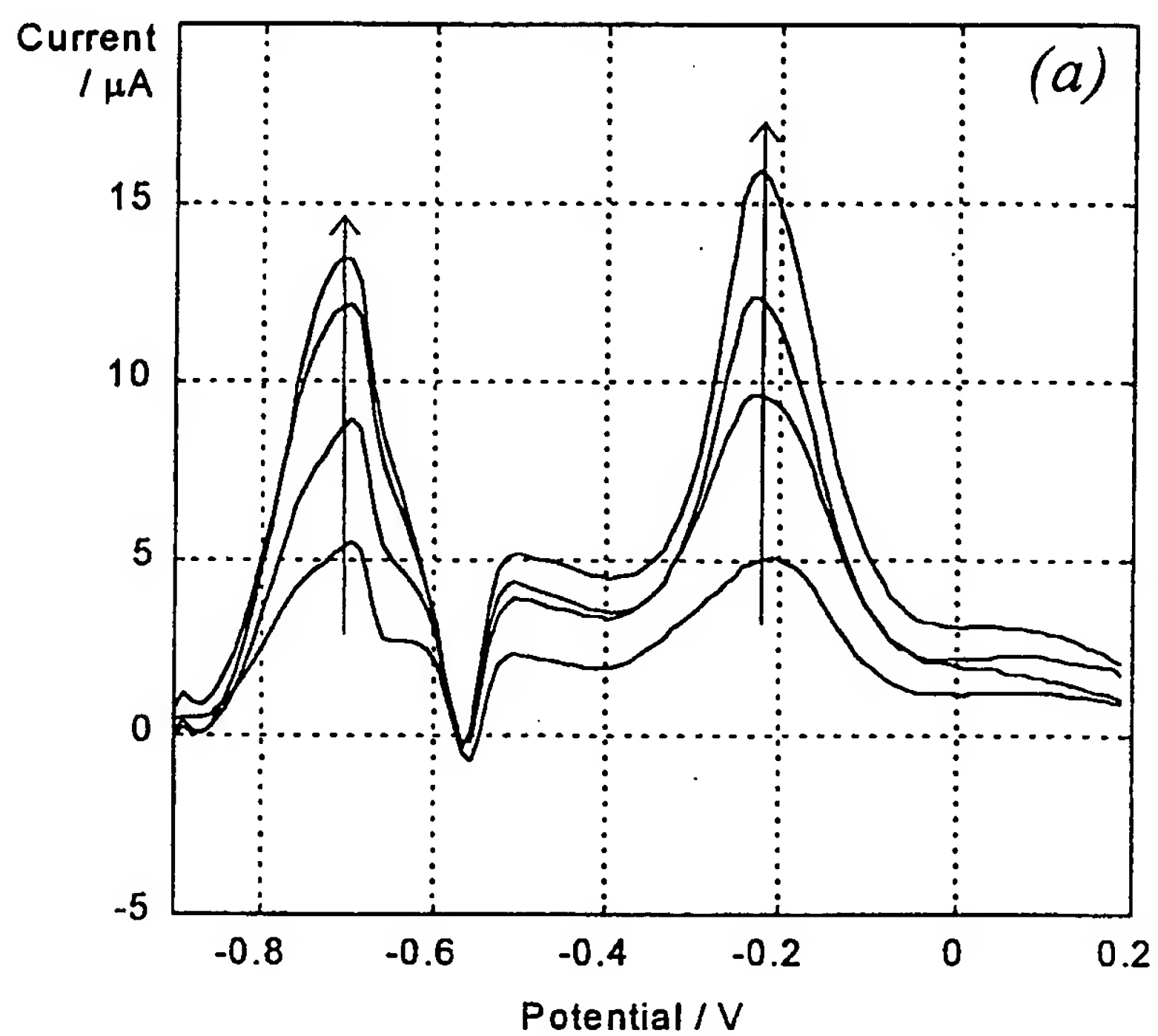


Fig 2 a,b

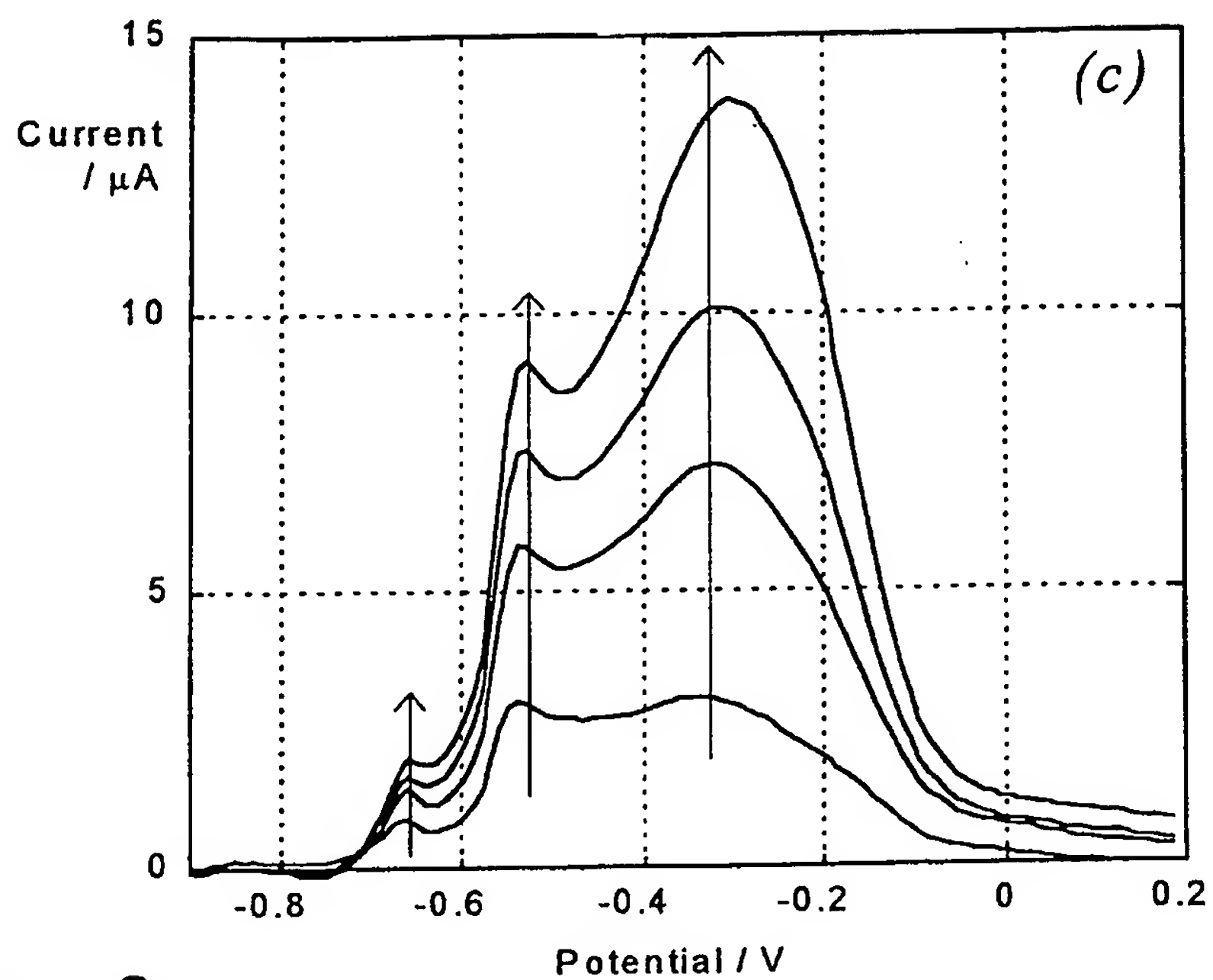


Fig 2c

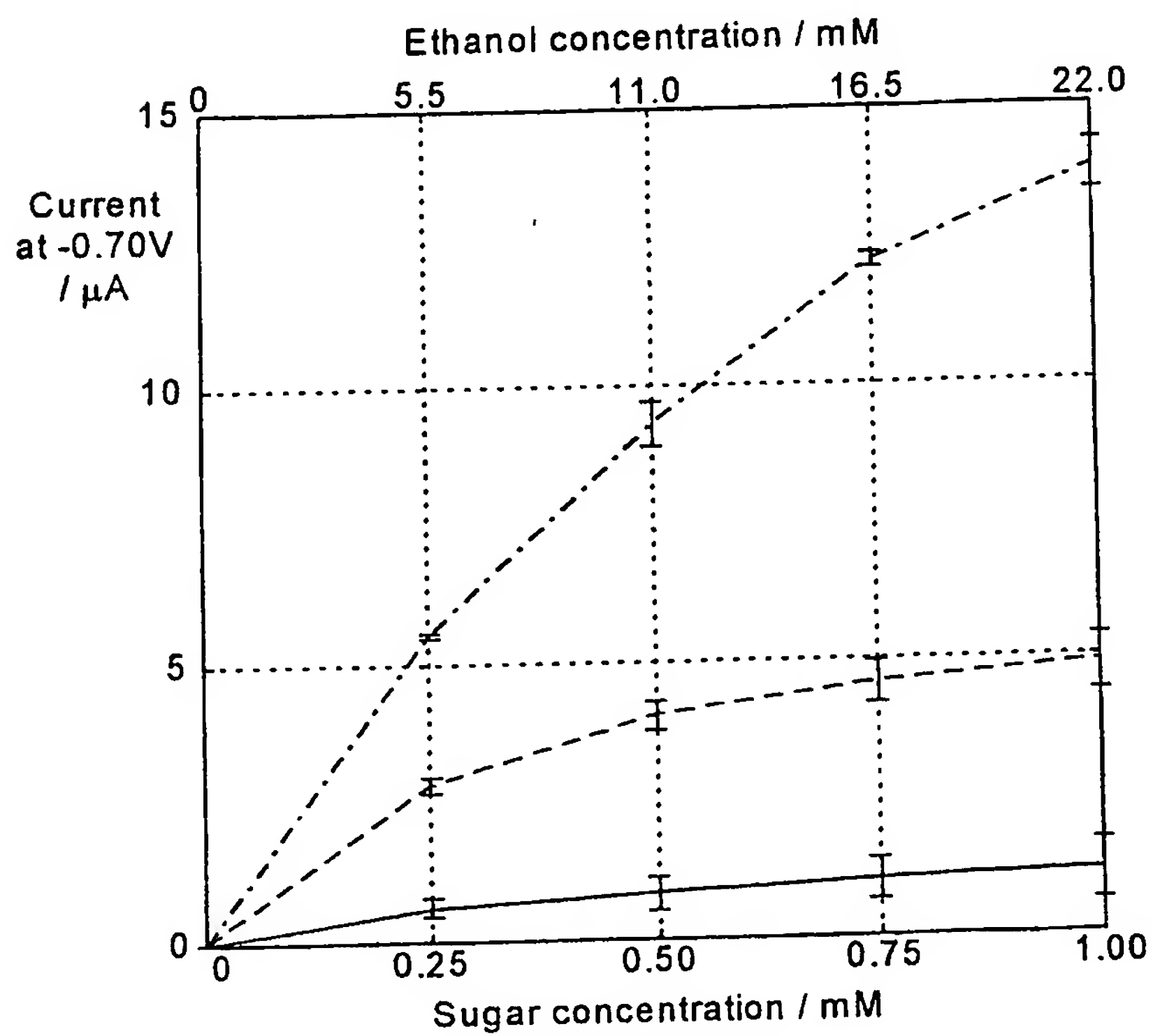


Fig 3a

Fig3(b)

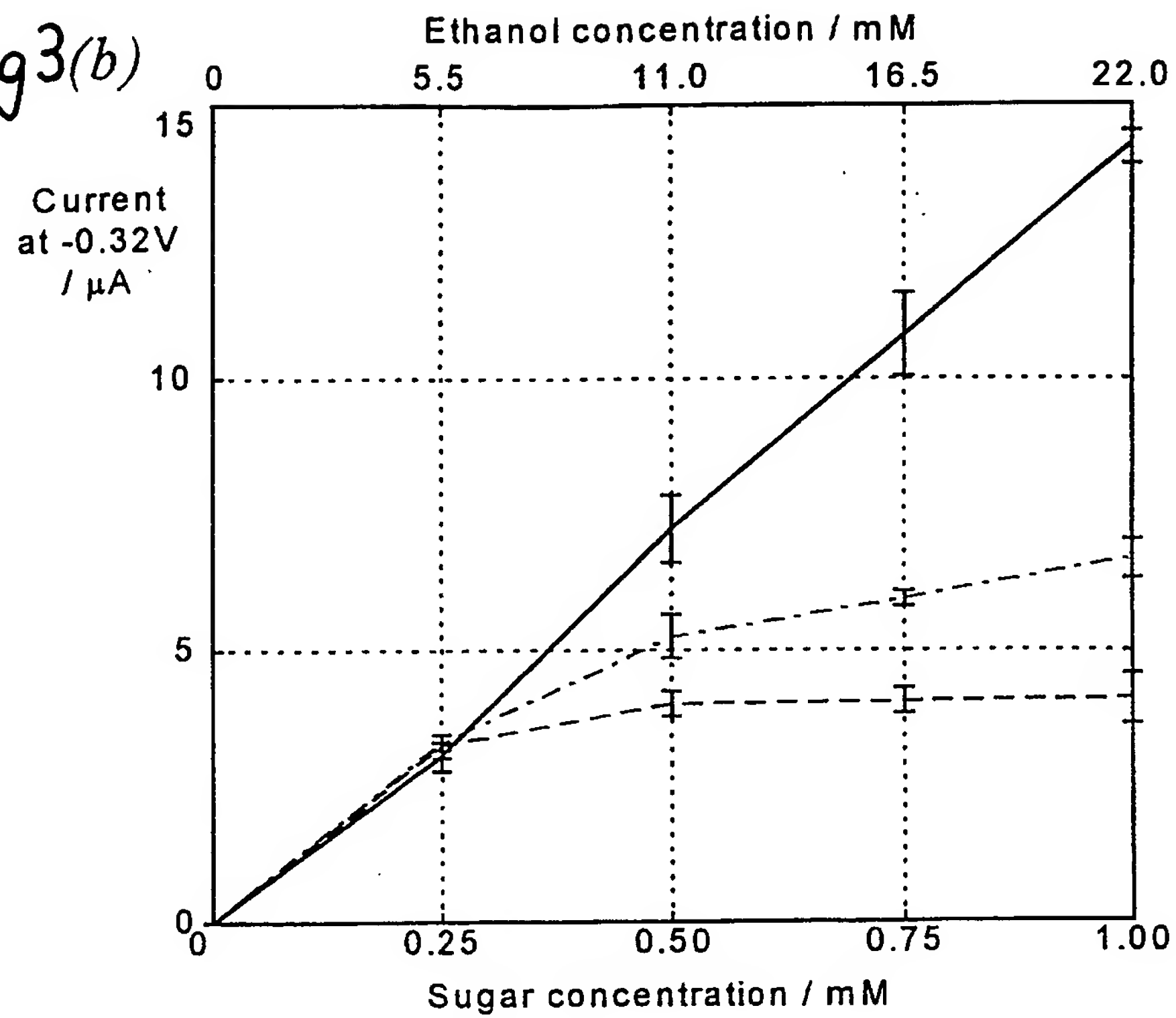
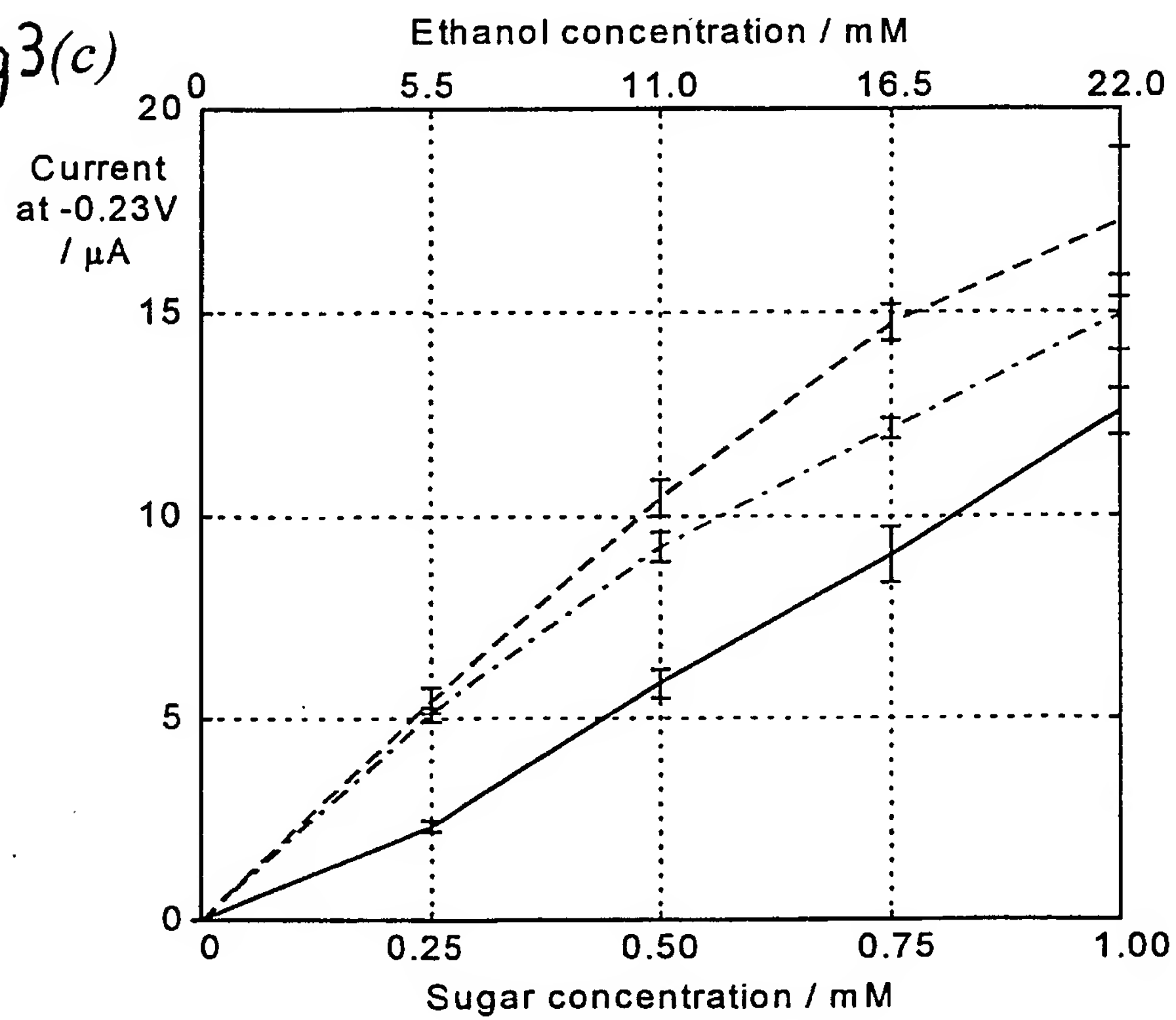


Fig3(c)



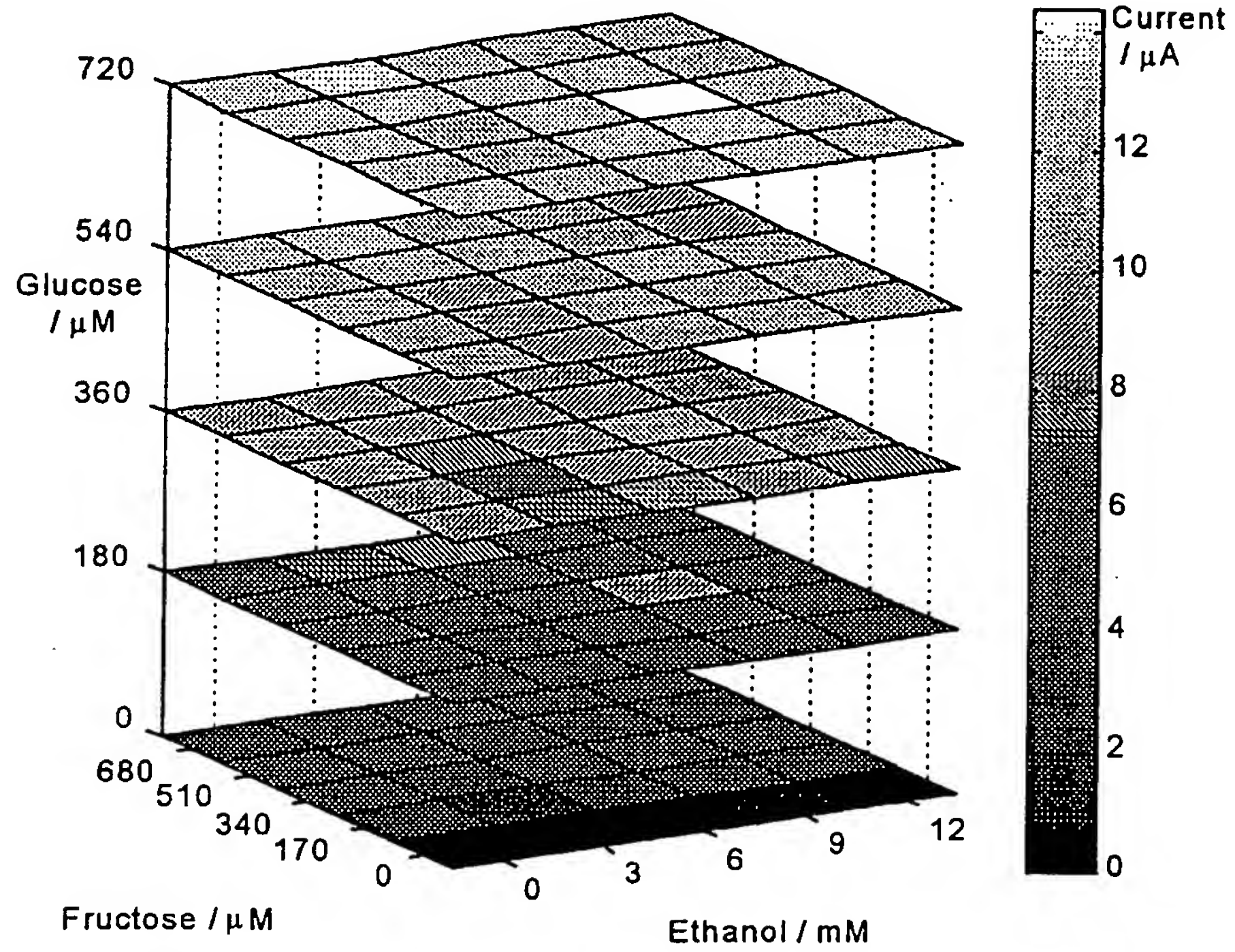


Fig 4a

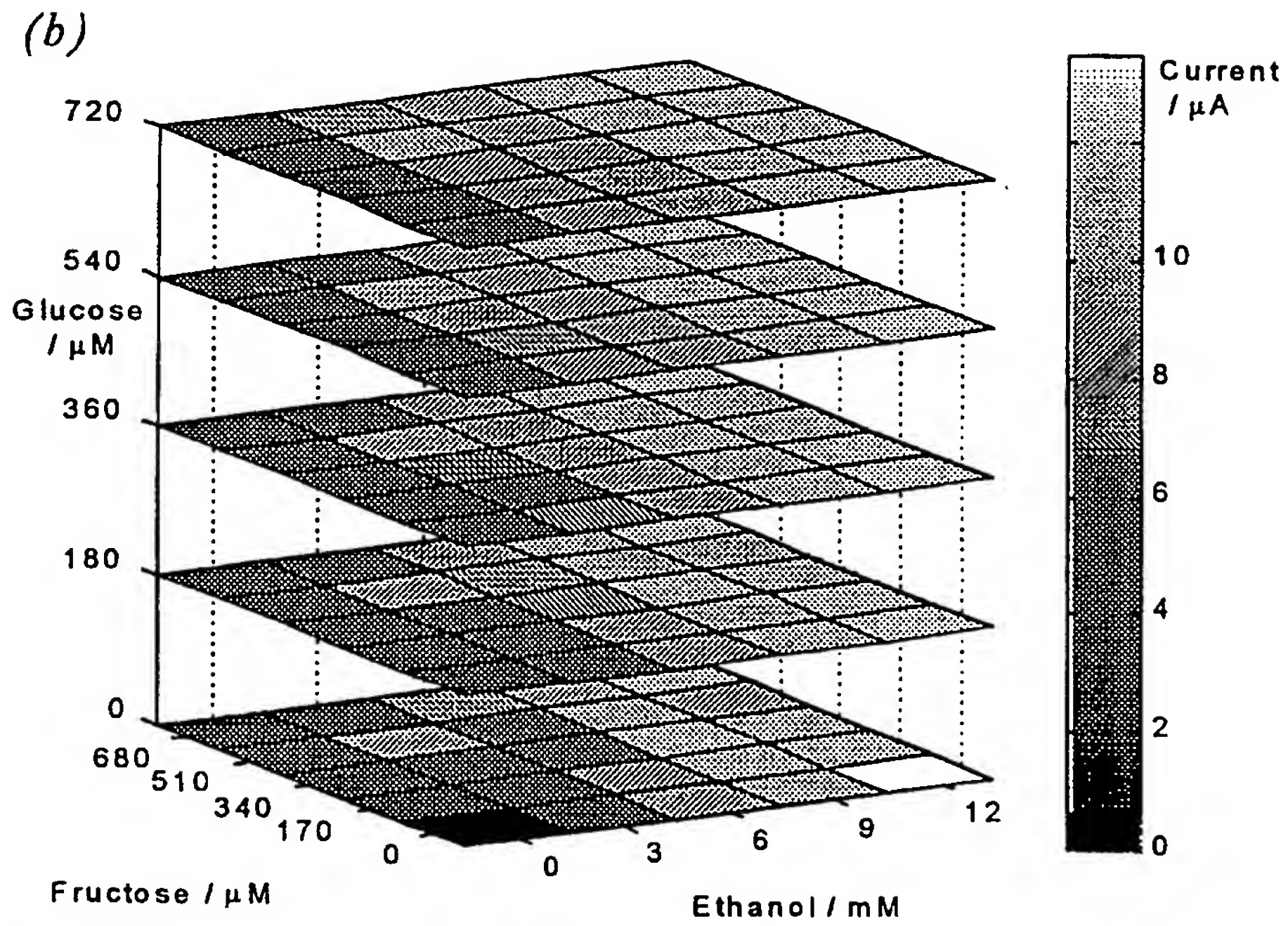


Fig 4b

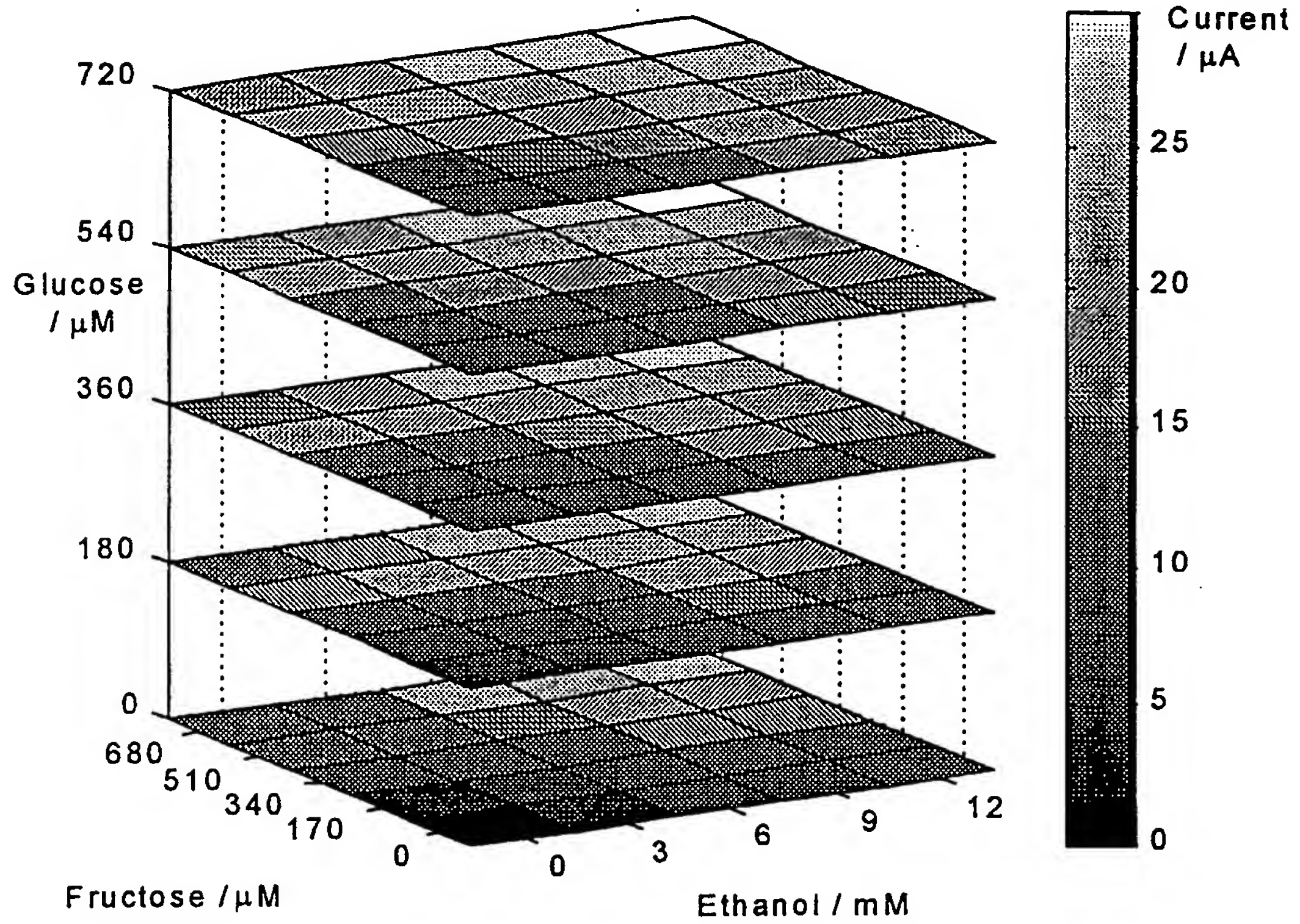


Fig 4c

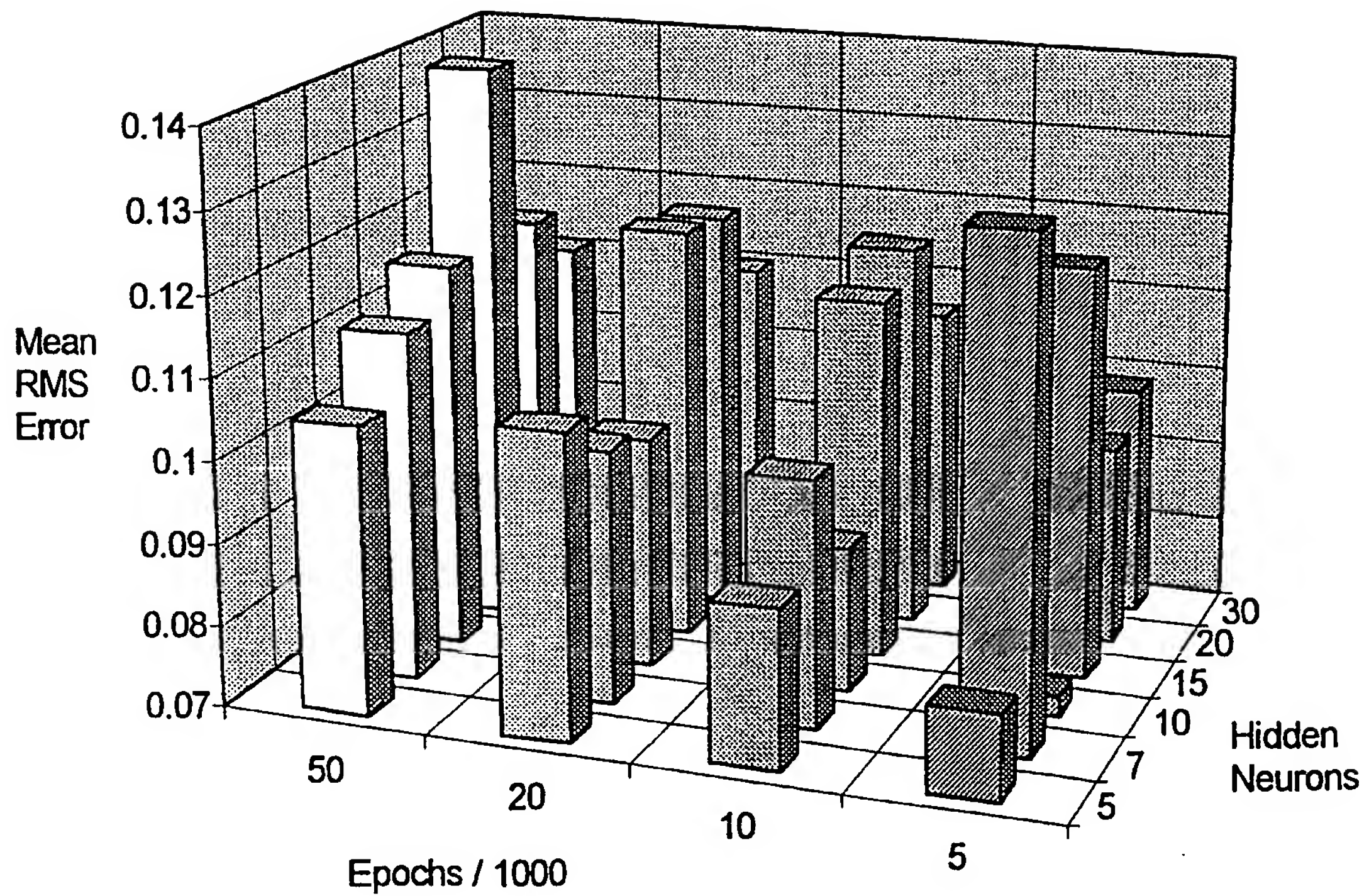


Fig 6

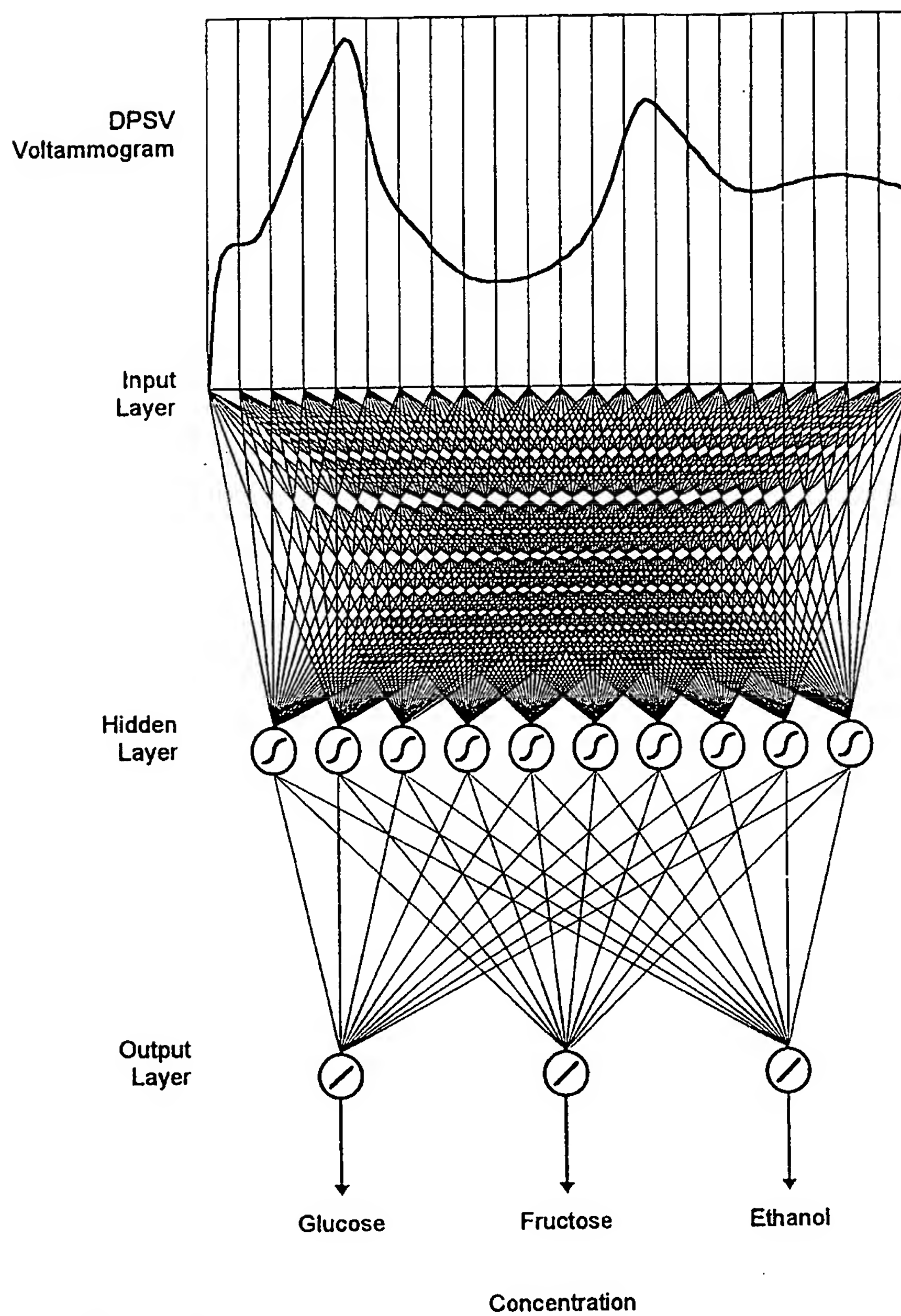


Fig 5

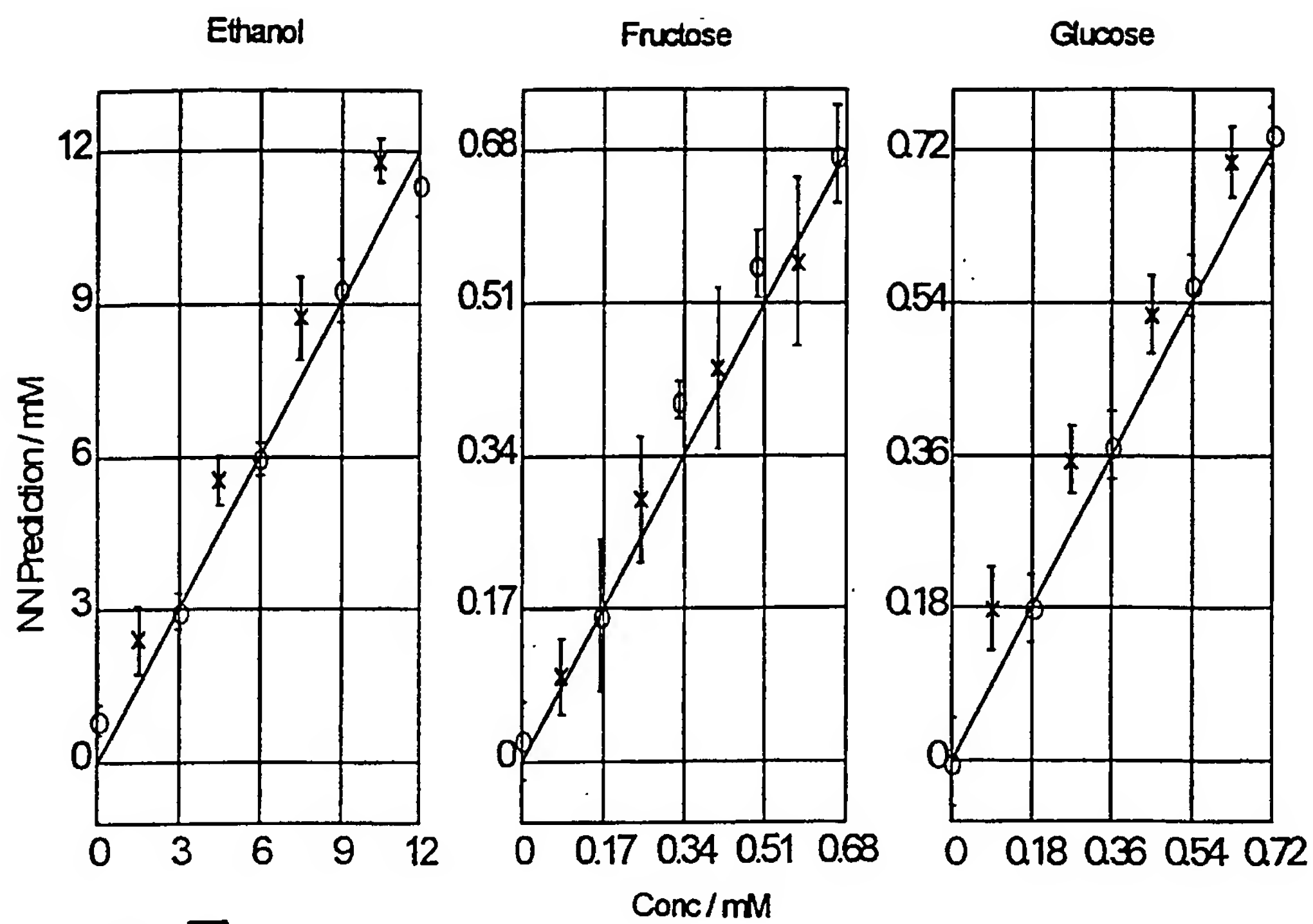


Fig 7

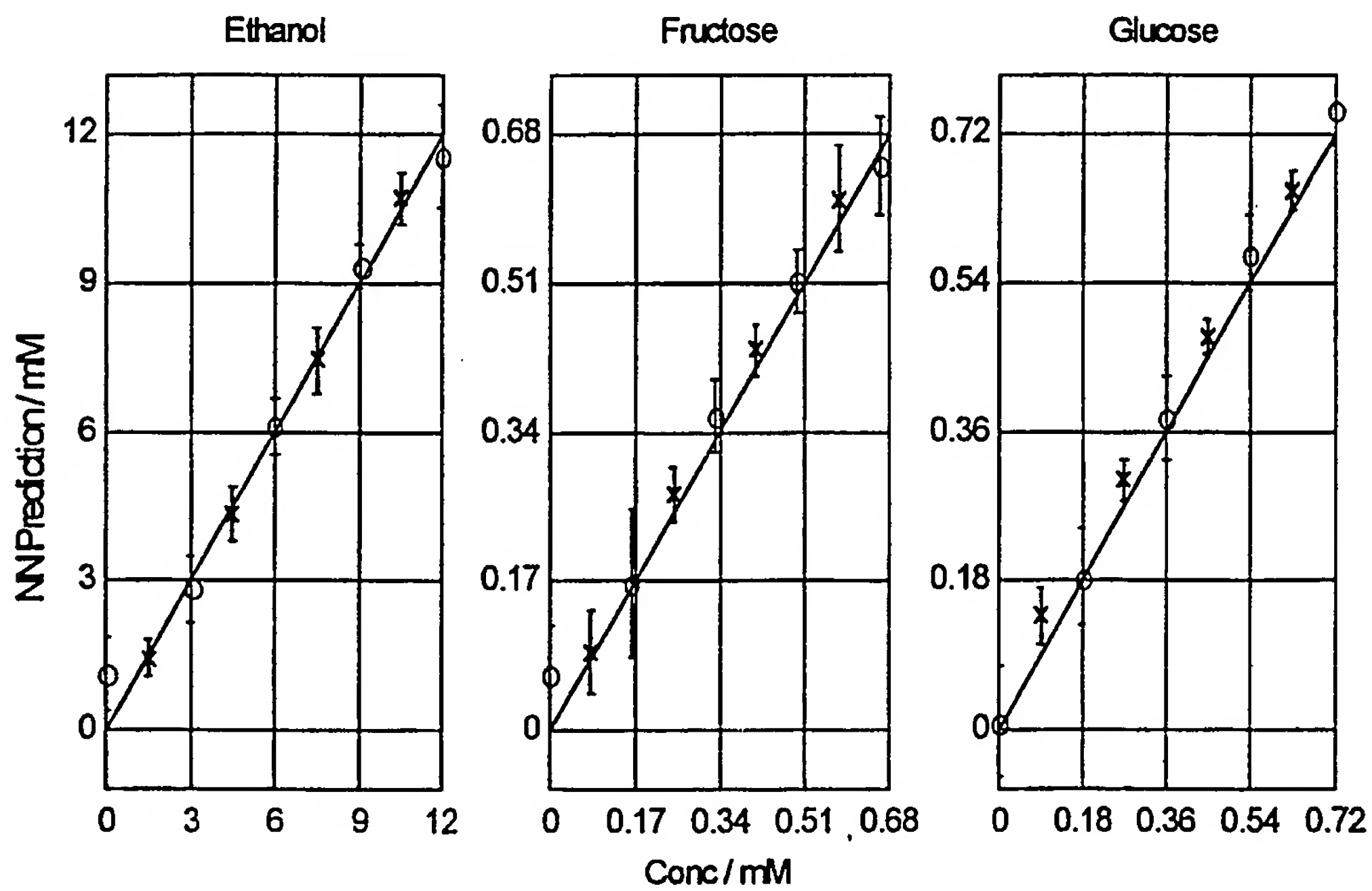


Fig 8

INTERNATIONAL SEARCH REPORT

International Application No.

PCT/GB 99/03263

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 G01N27/416

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 5 296 125 A (GLASS ROBERT S ET AL) 22 March 1994 (1994-03-22) column 7, line 38 -column 8, line 3 column 11, line 21 -column 12, line 34 column 14, line 29 - line 35 ---	1,19
Y	VOGT M ET AL: "A trainable cermet gas microsensor technology using cyclic voltammetry and neural networks" SENSORS AND ACTUATORS B,CH,ELSEVIER SEQUOIA S.A., LAUSANNE, vol. 36, no. 1, 1 October 1996 (1996-10-01), pages 370-376, XP004061097 ISSN: 0925-4005 the whole document --- -/--	1,19

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search

31 January 2000

Date of mailing of the international search report

14/02/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Duchatellier, M

INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 99/03263

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	VAN DER LINDEN: "arrays of electrodes for multicomponent analysis" ANALYTICAL PROCEEDINGS, vol. 26, no. 10, 1989, pages 329-331, XP002129224 gb the whole document ---	1
A	CH 659 327 A (INST ELEKTROKHEMIE AKADEMIE NA) 15 January 1987 (1987-01-15) abstract ---	1
A	US 5 554 273 A (DEMMIN HOLLIS C ET AL) 10 September 1996 (1996-09-10) abstract -----	1

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/GB 99/03263

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5296125 A	22-03-1994	US 5120421 A	09-06-1992
CH 659327 A	15-01-1987	NONE	
US 5554273 A	10-09-1996	EP 0756172 A	29-01-1997